



#6

OIEP

RAW SEQUENCE LISTING

DATE: 02/04/2002

PATENT APPLICATION: US/09/918,951

TIME: 11:21:56

Input Set : N:\Crf3\RULE60\09918951.raw

Output Set: N:\CRF3\02042002\I918951.raw

1 <110> APPLICANT: Alland, David
 2 Bloom, Barry R.
 3 Jacobs Jr., William R.
 4 <120> TITLE OF INVENTION: iniB, iniA AND iniC GENES OF MYCOBACTERIA AND METHODS
 5 OF USE
 6 <130> FILE REFERENCE: 96700/491
 7 <140> CURRENT APPLICATION NUMBER: 09/918,951
 8 <141> CURRENT FILING DATE: 2001-07-31
 10 <150> PRIOR APPLICATION NUMBER: US/09/177,349
 11 <151> PRIOR FILING DATE: 1998-10-23
 13 <160> NUMBER OF SEQ ID NOS: 14
 14 <170> SOFTWARE: PatentIn Ver. 2.0

ENTERED

16 <210> SEQ ID NO: 1
 17 <211> LENGTH: 53
 18 <212> TYPE: DNA
 19 <213> ORGANISM: Mycobacterium tuberculosis
 20 <400> SEQUENCE: 1
 21 cacggctacg acatccacgg ataagttccg gaccggcgta ggggtgcccc att 53
 23 <210> SEQ ID NO: 2
 24 <211> LENGTH: 5036
 25 <212> TYPE: DNA
 26 <213> ORGANISM: Mycobacterium tuberculosis
 27 <400> SEQUENCE: 2
 28 tccccctaatt ccctaaccg cgcgccaggc cgatcccgat aggtgtttgg ccggcttgcg 60
 29 gatcagaccc cgatttcggg gtgaggcgga atccatagcg tcgatggcac agcgccggtc 120
 30 acgcccggcg acagcttctt cgattgaagg gaaatgaaga tgacctcgct tatcgattac 180
 31 atcctgagcc tgttccgcag cgaagacgcc gcccggtcgt tcgttgccgc tccgggacgg 240
 32 gccatgacca gtgccgggct gatcgatatc gcgccgcacc aaatctcatc ggtggcgggc 300
 33 aatgtggtgc cgggtctgaa tctgggtgcc ggcgacccca tgagcggatt gcggcaggcc 360
 34 gtgcgcgctc ggcattggctt tgccgcaggac gtgcgcaatg tcggcttcgc cggtgacgcg 420
 35 ggcgcggggg tggcaagcgt catcacgacc gatgtcggtg cgggcctggc tagcggactg 480
 36 ggtgctgggt tcctgggtca gggtagcctg gctctcgccg cgtcaagcgg tggtttcggc 540
 37 ggtcaggctc gcttggtgc ccaggtcggt ctgggtttta ctgccgtgat tgaggccgag 600
 38 gtgcgcgctc aggttgggtc tgggttaggt attgggacgg gtctgggtgc tcaggccggg 660
 39 atgggctttg gcggcggggt tggcctgggt ctgggtggtc aggcggcgcg tgtgatcggt 720
 40 gggagcgcgg ccggggctat cgggtgccggc gtccggcggt gcctaggcgg caatggccag 780
 41 atcggagttg ccggccaggg tgccgttgcc gctgggtgct gcgctggtgt cggcgggccag 840
 42 gcgggcacgc ctaccagat cgggtgtctca gccgggtggg ggctcggcgg cgtcgccaat 900
 43 gtccagcgcc tgaccggggg cagcagcaac cgagtgttgg cttccaacgc aagcgggccag 960
 44 gcgggggttga tcgccagtga aggcgctgcc ttgaacggcg ctgctatgcc tcatctgtcg 1020
 45 ggcccgttag ccggtgtcgg tgtgggtggg caggccggcg ccgctggcgg cgccgggttg 1080
 46 ggcttcggag cggtcgggca cccgactcct cagccggcgg ccctggggcg ggctggcggtg 1140
 47 gtggccaaga ccgaggcgcc tgetggagtg gttggcgggg tcggcggggc aaccgcggcc 1200

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/918,951

DATE: 02/04/2002

TIME: 11:21:56

Input Set : N:\Crif3\RULE60\09918951.raw

Output Set: N:\CRF3\02042002\I918951.raw

48	ggggtcggcg	gggcacacgg	cgacatcctg	ggccacgagg	gagccgcact	gggcagtgtc	1260
49	gacacgggtca	acgccgggtgt	cacgcccgtc	gagcatggct	tggtcctgcc	cagtggcccc	1320
50	ctgatccacg	gcggtaccgg	cggctatggc	ggcatgaacc	cgccagtga	cgatgcggcg	1380
51	gcaccgcaag	ttccggcgcg	ggcccagccg	atgaccacgg	cggccgagca	cacgccggcg	1440
52	gttaccacaac	cgcagcacac	gccggtcgag	ccgccgggtcc	acgataagcc	gccgagccat	1500
53	tcggtgtttg	acgtcgggtca	cgagccggcg	gtgacgcaca	cgccgcccgc	gcccacgaa	1560
54	ctgccgtcgt	acggcctttt	cggactaccc	gggttctgat	tcgcgagccg	atttcacgaa	1620
55	ccggtgggga	cgttcattgt	ccccgccgtt	ttgtgcgcat	accgtgatct	gaggcgtaaa	1680
56	cgagcgagaa	agtggggcga	caagggtgacc	cagcccgatg	acccacgtcg	ggtcggtgtg	1740
57	atcgtcgaac	tgatcgatca	cactatcgcc	atcgccaaac	tgaacgagcg	tggtgatcta	1800
58	gtacagcggg	tgacgcgggc	tcgccagcgg	atcaccgacc	cgcagggtccg	tgtggtgatc	1860
59	gccgggctgc	tcaaacaggg	caagagtcaa	ttgtctcaatt	cgttgcctcaa	cctgcccgcg	1920
60	gcgcgagtag	gcgatgacga	ggccaccgtg	gtgatcaccg	tcgtaagcta	cagcgcccaa	1980
61	ccgtcggccc	ggcttgtgct	ggccgcgggg	cccgcagggga	caaccgcagc	ggttgacatt	2040
62	cccgtcgatg	acatcagcac	cgatgtgcgt	cgggctccgc	acgccggtgg	ccgcgaggtg	2100
63	ttgcgggtcg	aggtcggcgc	gccagcccg	ctgctgcggg	gcgggctggc	gtttatcgat	2160
64	actccgggtg	tgggcggcct	cggacagccc	cacctgtcgg	cgacgctggg	gctgctaccc	2220
65	gaggccgatg	ccgtcttggg	ggtcagcgac	accagccagg	aattcaccga	accgcagatg	2280
66	tggttcgtgc	ggcaggccca	ccagatctgt	ccggtcgggg	cggtcgtggc	caccaagacc	2340
67	gacctgtatc	cgcgctggcg	ggagatcgtc	aatgccaatg	cagcacatct	gcagcggggc	2400
68	cgggttccga	tgccgatcat	cgcagttctca	tactgttgcc	gcagccacgc	ggtcacgctt	2460
69	aacgacaaa	agctcaacga	agagtccaac	tttccggcga	tcgtcaagtt	tctcagcgag	2520
70	caggtgcttt	cccgcgcgac	ggagcgagtg	cgtgctgggg	tactcggcga	aatacgttcg	2580
71	gcaacagagc	aattggcggt	gtctctaggt	tccgaactat	cgttggtcaa	cgaccgcaac	2640
72	ctccgtgacc	gacttgcttc	ggatttggag	cggcgcaaac	gggaagccca	gcaggcggtg	2700
73	caacagacag	cgtgtgggca	gcaggtgctg	ggcgacgggt	tcaacgacct	gactgctgac	2760
74	gtggaccaac	acctacgaac	ccgcttcggc	accgtcaccg	aagacgcccga	gcgccagatc	2820
75	gactcctgtg	acccgactgc	gcattggggc	gagattggca	acgacgtcga	gaatgcgac	2880
76	gccacagcgg	tcggcgacaa	cttcgtgtgg	gcataccagc	gttccgaagc	gttggccgac	2940
77	gacgtcgetc	gtccctttgc	cgacgcgggg	ttggactcgg	tcctgtcagc	agagctgagc	3000
78	ccccacgtca	tgggcaccga	cttcggccgg	ctcaaagcgc	tgggccggat	ggaatcgaaa	3060
79	ccgctgcgcc	ggggccataa	aatgattatc	ggcatgcggg	gttcctatgg	cggcgtggtc	3120
80	atgattggca	tgctgtcgtc	ggtggtcgga	cttgggttgt	tcaacccgct	atcggtgggg	3180
81	gccgggttga	tcctcggccg	gatggcatat	aaagaggaca	aacaaaaccg	gttgctgcgg	3240
82	gtgcgcagcg	aggccaaggc	caatgtgcgg	cgcttcgtcg	acgacatttc	gttcgtcgtc	3300
83	agcaaaacaat	cacgggatcg	gtcgaagatg	atccagcgtc	tgctgcgcga	ccactaccgc	3360
84	gagatcgccg	aagagatcac	ccggtcgcct	accgagtcct	tgacggcgac	catcgcggcg	3420
85	gcgcaggtgg	cggaaaaccga	gcgggacaat	cgaattcggg	aaacttcagcg	gcaattgggt	3480
86	atcctgagcc	aggtcaacga	caaccttgcc	ggcttggagc	caaccttgac	gccccggggc	3540
87	agcttgggac	gagcgtgagc	accagcgacc	gggtccgcgc	gattctgcac	gcaaccatcc	3600
88	aggcctaccg	gggtgcgcgg	gcctatcgtc	agcgtggcga	cgttttttgc	cagctggacc	3660
89	gcacggtg	gcgcctagcc	gaaccgctgc	gcacgcggtt	ggctggcaca	ctcaaggccg	3720
90	gaaaatccac	tctcgtcaac	gcccttgcgc	gcgacgacat	cgtcccgacc	gatgccaccg	3780
91	aggccaaccg	gattgtgacc	tggttcggcg	accgtccgac	accgcgggtc	accgccaacc	3840
92	atcgcgcgcg	tcgacgcgcc	aacgtgcgca	tcacccgctc	gggcgggctg	agtttcgacc	3900
93	tgcgagggat	caaccgcggc	gagctgatcg	acctggaagt	cgagtggcca	gccgaggaac	3960
94	tcacgcagcg	caccattgtt	gacaccccg	gaacgtcgtc	gttggcatgc	gatgcctccg	4020
95	agcgcacgtt	gcggctgctg	gtccccgccg	acgggggtgc	tcgggtggat	gcgggtggtg	4080
96	tcctgttgcg	caccctgaac	gccgctgacg	tcgcgctgct	caaacagatc	ggtgggctgg	4140

RAW SEQUENCE LISTING

DATE: 02/04/2002

PATENT APPLICATION: US/09/918,951

TIME: 11:21:56

Input Set : N:\Crf3\RULE60\09918951.raw

Output Set: N:\CRF3\02042002\I918951.raw

```

97      tcggcggggtc ggtgggagcc ctgggcatca tcgggggtggc gtctcgcgcg gatgagatcg 4200
98      gcgcggggccg catcgacgcg atgctctcgg ccaacgacgt ggccaagcgg ttcacccgcg 4260
99      aactgaacca gatgggcatt tgccaggcgg tgggtgccgt atccggactt cttgcgctga 4320
100     ccgcgcgcac actgcgccag accgagttca tcgcgctgcg caagctggcc ggtgccgagc 4380
101     gcaccgagct caatagggcc ctgctgagcg tggaccgttt tgtgcgccgg gacagtccgc 4440
102     taccggtgga cgcgggcatc cgtgcgcaat tgctcgagcg gttcggcatg ttcggcatcc 4500
103     ggatgtcgat tgccgtgctg gcggccggcg tgaccgattc gaccgggctg gccgccgaac 4560
104     tgctggagcg cagcgggctg gtggcgctgc gcaatgtgat agaccagcag ttcgcgcagc 4620
105     gctccgacat gcttaaggcg cataccgctt tggctctcct gcgccgattc gtgcagacgc 4680
106     atccggtgcc ggccgaccccg tacgtcattg ccgacatcga ccggttgcta gccgacaccc 4740
107     acgccttcga agaactccga atgctaagcc ttttgcttc gcgggcaacg acattgaacg 4800
108     acgacgaaat cgcgtcgctg cgcgcgcatc tcggcgggtc gggcaccagt gccgccgctc 4860
109     ggctgggcct ggatccccgcg aattctcgcg aggccccgcg cgcgcgctg gccgcagcgc 4920
110     aacactggcg tcgccgtgcg gcgcattccac tcaacgatcc gttcactacc agggcctgtc 4980
111     gcgcggcggt gcgcagcgcc gaggcgatgg tggcgagatt ctctgctcgc cgctga      5036

```

113 <210> SEQ ID NO: 3

114 <211> LENGTH: 479

115 <212> TYPE: PRT

116 <213> ORGANISM: Mycobacterium tuberculosis

117 <400> SEQUENCE: 3

```

118      Met Thr Ser Leu Ile Asp Tyr Ile Leu Ser Leu Phe Arg Ser Glu Asp
119           1           5           10           15
120      Ala Ala Arg Ser Phe Val Ala Ala Pro Gly Arg Ala Met Thr Ser Ala
121           20           25           30
122      Gly Leu Ile Asp Ile Ala Pro His Gln Ile Ser Ser Val Ala Ala Asn
123           35           40           45
124      Val Val Pro Gly Leu Asn Leu Gly Ala Gly Asp Pro Met Ser Gly Leu
125           50           55           60
126      Arg Gln Ala Val Ala Ala Arg His Gly Phe Ala Gln Asp Val Ala Asn
127           65           70           75           80
128      Val Gly Phe Ala Gly Asp Ala Gly Ala Gly Val Ala Ser Val Ile Thr
129           85           90           95
130      Thr Asp Val Gly Ala Gly Leu Ala Ser Gly Leu Gly Ala Gly Phe Leu
131           100          105          110
132      Gly Gln Gly Gly Leu Ala Leu Ala Ala Ser Ser Gly Gly Phe Gly Gly
133           115          120          125
134      Gln Val Gly Leu Ala Ala Gln Val Gly Leu Gly Phe Thr Ala Val Ile
135           130          135          140
136      Glu Ala Glu Val Gly Ala Gln Val Gly Ala Gly Leu Gly Ile Gly Thr
137           145          150          155          160
138      Gly Leu Gly Ala Gln Ala Gly Met Gly Phe Gly Gly Gly Val Gly Leu
139           165          170          175
140      Gly Leu Gly Gly Gln Ala Gly Gly Val Ile Gly Gly Ser Ala Ala Gly
141           180          185          190
142      Ala Ile Gly Ala Gly Val Gly Gly Arg Leu Gly Gly Asn Gly Gln Ile
143           195          200          205
144      Gly Val Ala Gly Gln Gly Ala Val Gly Ala Gly Val Gly Ala Gly Val
145           210          215          220
146      Gly Gly Gln Ala Gly Ile Ala Ser Gln Ile Gly Val Ser Ala Gly Gly

```

RAW SEQUENCE LISTING

DATE: 02/04/2002

PATENT APPLICATION: US/09/918,951

TIME: 11:21:56

Input Set : N:\Crif3\RULE60\09918951.raw

Output Set: N:\CRF3\02042002\I918951.raw

```

147      225      230      235      240
148 Gly Leu Gly Gly Val Gly Asn Val Ser Gly Leu Thr Gly Val Ser Ser
149      245      250      255
150 Asn Ala Val Leu Ala Ser Asn Ala Ser Gly Gln Ala Gly Leu Ile Ala
151      260      265      270
152 Ser Glu Gly Ala Ala Leu Asn Gly Ala Ala Met Pro His Leu Ser Gly
153      275      280      285
154 Pro Leu Ala Gly Val Gly Val Gly Gly Gln Ala Gly Ala Ala Gly Gly
155      290      295      300
156 Ala Gly Leu Gly Phe Gly Ala Val Gly His Pro Thr Pro Gln Pro Ala
157      305      310      315      320
158 Ala Leu Gly Ala Ala Gly Val Val Ala Lys Thr Glu Ala Ala Ala Gly
159      325      330      335
160 Val Val Gly Gly Val Gly Gly Ala Thr Ala Ala Gly Val Gly Gly Ala
161      340      345      350
162 His Gly Asp Ile Leu Gly His Glu Gly Ala Ala Leu Gly Ser Val Asp
163      355      360      365
164 Thr Val Asn Ala Gly Val Thr Pro Val Glu His Gly Leu Val Leu Pro
165      370      375      380
166 Ser Gly Pro Leu Ile His Gly Gly Thr Gly Gly Tyr Gly Gly Met Asn
167      385      390      395      400
168 Pro Pro Val Thr Asp Ala Pro Ala Pro Gln Val Pro Ala Arg Ala Gln
169      405      410      415
170 Pro Met Thr Thr Ala Ala Glu His Thr Pro Ala Val Thr Gln Pro Gln
171      420      425      430
172 His Thr Pro Val Glu Pro Pro Val His Asp Lys Pro Pro Ser His Ser
173      435      440      445
174 Val Phe Asp Val Gly His Glu Pro Pro Val Thr His Thr Pro Pro Ala
175      450      455      460
176 Pro Ile Glu Leu Pro Ser Tyr Gly Leu Phe Gly Leu Pro Gly Phe
177      465      470      475
179 <210> SEQ ID NO: 4
180 <211> LENGTH: 640
181 <212> TYPE: PRT
182 <213> ORGANISM: Mycobacterium tuberculosis
183 <400> SEQUENCE: 4
184 Met Val Pro Ala Gly Leu Cys Ala Tyr Arg Asp Leu Arg Arg Lys Arg
185      1      5      10      15
186 Ala Arg Lys Trp Gly Asp Thr Val Thr Gln Pro Asp Asp Pro Arg Arg
187      20      25      30
188 Val Gly Val Ile Val Glu Leu Ile Asp His Thr Ile Ala Ile Ala Lys
189      35      40      45
190 Leu Asn Glu Arg Gly Asp Leu Val Gln Arg Leu Thr Arg Ala Arg Gln
191      50      55      60
192 Arg Ile Thr Asp Pro Gln Val Arg Val Val Ile Ala Gly Leu Leu Lys
193      65      70      75      80
194 Gln Gly Lys Ser Gln Leu Leu Asn Ser Leu Leu Asn Leu Pro Ala Ala
195      85      90      95
196 Arg Val Gly Asp Asp Glu Ala Thr Val Val Ile Thr Val Val Ser Tyr

```

RAW SEQUENCE LISTING

DATE: 02/04/2002

PATENT APPLICATION: US/09/918,951

TIME: 11:21:56

Input Set : N:\Crf3\RULE60\09918951.raw

Output Set: N:\CRF3\02042002\I918951.raw

197		100		105		110
198	Ser Ala Gln Pro Ser Ala Arg Leu Val Leu Ala Ala Gly Pro Asp Gly					
199		115		120		125
200	Thr Thr Ala Ala Val Asp Ile Pro Val Asp Asp Ile Ser Thr Asp Val					
201		130		135		140
202	Arg Arg Ala Pro His Ala Gly Gly Arg Glu Val Leu Arg Val Glu Val					
203	145		150		155	160
204	Gly Ala Pro Ser Pro Leu Leu Arg Gly Gly Leu Ala Phe Ile Asp Thr					
205		165		170		175
206	Pro Gly Val Gly Gly Leu Gly Gln Pro His Leu Ser Ala Thr Leu Gly					
207		180		185		190
208	Leu Leu Pro Glu Ala Asp Ala Val Leu Val Val Ser Asp Thr Ser Gln					
209		195		200		205
210	Glu Phe Thr Glu Pro Glu Met Trp Phe Val Arg Gln Ala His Gln Ile					
211		210		215		220
212	Cys Pro Val Gly Ala Val Val Ala Thr Lys Thr Asp Leu Tyr Pro Arg					
213	225		230		235	240
214	Trp Arg Glu Ile Val Asn Ala Asn Ala Ala His Leu Gln Arg Ala Arg					
215		245		250		255
216	Val Pro Met Pro Ile Ile Ala Val Ser Ser Leu Leu Arg Ser His Ala					
217		260		265		270
218	Val Thr Leu Asn Asp Lys Glu Leu Asn Glu Glu Ser Asn Phe Pro Ala					
219		275		280		285
220	Ile Val Lys Phe Leu Ser Glu Gln Val Leu Ser Arg Ala Thr Glu Arg					
221		290		295		300
222	Val Arg Ala Gly Val Leu Gly Glu Ile Arg Ser Ala Thr Glu Gln Leu					
223	305		310		315	320
224	Ala Val Ser Leu Gly Ser Glu Leu Ser Val Val Asn Asp Pro Asn Leu					
225		325		330		335
226	Arg Asp Arg Leu Ala Ser Asp Leu Glu Arg Arg Lys Arg Glu Ala Gln					
227		340		345		350
228	Gln Ala Val Gln Gln Thr Ala Leu Trp Gln Gln Val Leu Gly Asp Gly					
229		355		360		365
230	Phe Asn Asp Leu Thr Ala Asp Val Asp His Asp Leu Arg Thr Arg Phe					
231		370		375		380
232	Arg Thr Val Thr Glu Asp Ala Glu Arg Gln Ile Asp Ser Cys Asp Pro					
233	385		390		395	400
234	Thr Ala His Trp Ala Glu Ile Gly Asn Asp Val Glu Asn Ala Ile Ala					
235		405		410		415
236	Thr Ala Val Gly Asp Asn Phe Val Trp Ala Tyr Gln Arg Ser Glu Ala					
237		420		425		430
238	Leu Ala Asp Asp Val Ala Arg Ser Phe Ala Asp Ala Gly Leu Asp Ser					
239		435		440		445
240	Val Leu Ser Ala Glu Leu Ser Pro His Val Met Gly Thr Asp Phe Gly					
241		450		455		460
242	Arg Leu Lys Ala Leu Gly Arg Met Glu Ser Lys Pro Leu Arg Arg Gly					
243	465		470		475	480
244	His Lys Met Ile Ile Gly Met Arg Gly Ser Tyr Gly Gly Val Val Met					
245		485		490		495

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/918,951

DATE: 02/04/2002

TIME: 11:21:57

Input Set : N:\Crf3\RULE60\09918951.raw

Output Set: N:\CRF3\02042002\I918951.raw